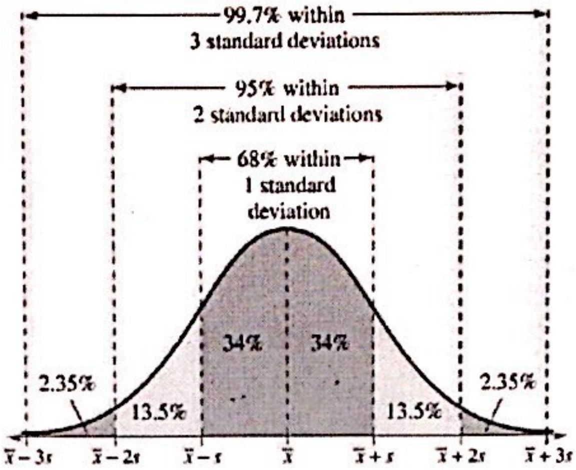
Notes – 2.4 Empirical Rule

Many real-life data sets have distributions that are approximately symmetric and bell-shaped. For instance, the distributions of men's and women's heights in the United States are approximately symmetric and bell-shaped (see the figures at the left and bottom left). Later in the text, you will study bell-shaped distributions in greater detail. For now, however, the **Empirical Rule** can help you see how valuable the standard deviation can be as a measure of variation.

Bell-Shaped Distribution



EMPIRICAL RULE (OR 68-95-99.7 RULE)

For data sets with distributions that are approximately symmetric and bell-shaped, the standard deviation has these characteristics.

1. About 68% of the data lie within one standard deviation of the mean.
2. About 95% of the data lie within two standard deviations of the mean.
3. About 99.7% of the data lie within three standard deviations of the mean.

EXAMPLE 6

Using the Empirical Rule

In a survey conducted by the National Center for Health Statistics, the sample mean height of women in the United States (ages 20–29) was 64.2 inches, with a sample standard deviation of 2.9 inches. Estimate the percent of women whose heights are between 58.4 inches and 64.2 inches.

Try It Yourself 6

Estimate the percent of women ages 20–29 whose heights are between 64.2 inches and 67.1 inches.

a. How many standard deviations is 67.1 to the right of 64.2?

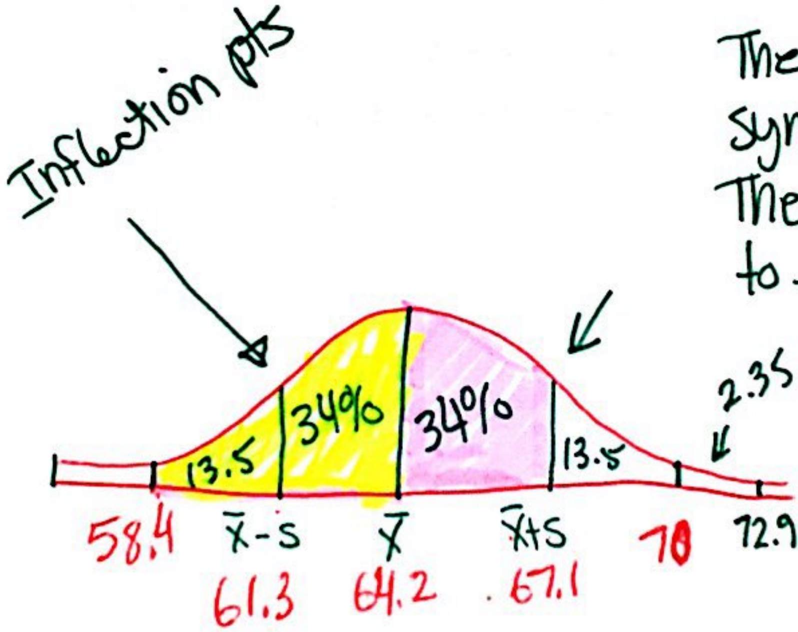
b. Use the Empirical Rule to estimate the percent of the data between 64.2 and 67.1.

c. Interpret the result in the context of the data.

The normal curve is determined by the mean, $\bar{x}$, and the standard deviation s (σ).

The normal curve is symmetric about $\bar{x}$.
The tails are asymptotic to the horizontal axis.

Ex 6



$$\begin{array}{r} 64.2 \\ +2.9 \\ \hline 67.1 \\ +2.9 \\ \hline 70.0 \end{array}$$

$x < X <$

58.4" 64.2"

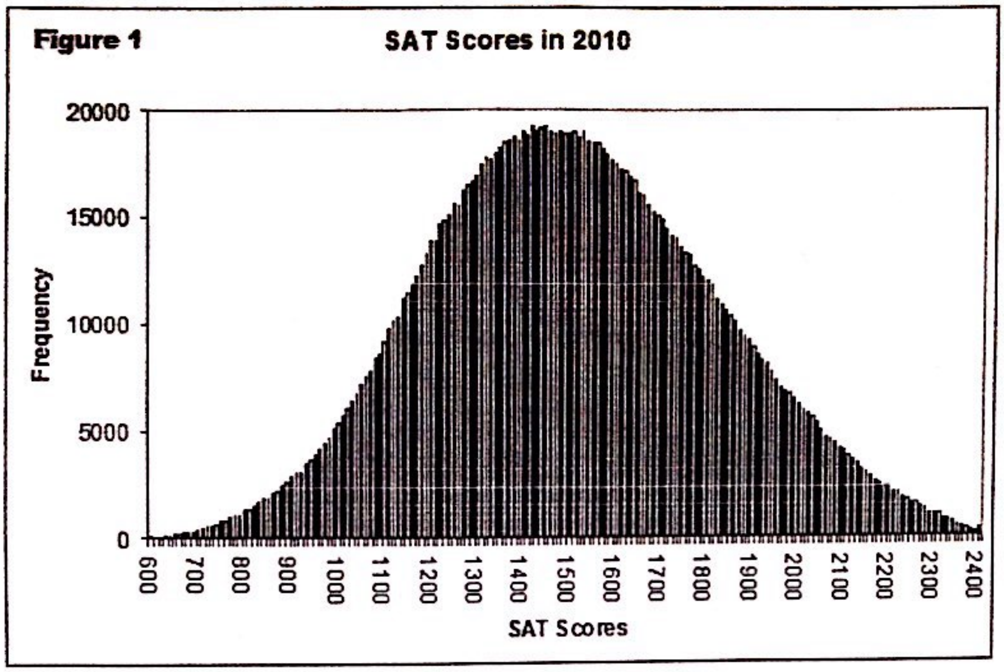
$13.5 + 34 = 47.5\%$

$$\begin{array}{r} 64.2 \\ -2.9 \\ \hline 61.3 \\ -2.9 \\ \hline 58.4 \end{array}$$

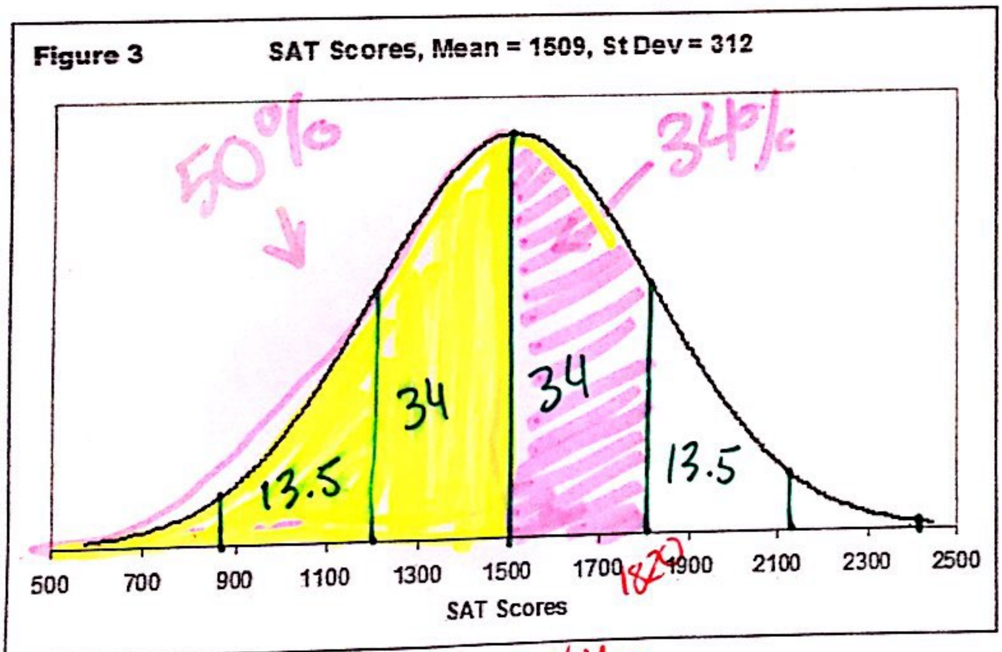
Try It Yourself 6

34%

Figure 1 below is a histogram of all 1,547,990 SAT scores taken in 2010.



There are 181 bars in this histogram, corresponding to the 181 distinct possible scores in the data (the scores come with an increment of 10, ranging from 600, 610, 620, and all the way to 2390, and 2400). Because of many bars crammed into a small graph, each bar appears as a thin vertical line. The histogram is symmetrical around a single peak and it tapers down smoothly on each side. Most of the data is clustered in the middle. The bars around the middle are very tall (i.e. most students score in the middle range). On the other hand, the bars at either the left side or the right side are very short (very few students score at the top and at the bottom). As a result, the histogram has a "bell" shape. The following (Figure 2) is a representation of the same 1,547,990 SAT scores as a smooth curve, which also has a "bell" shape.



$1509 = \bar{x}$
$+312$
$1821 = \bar{x} + s$
$+312$
$2133 = \bar{x} + 2s$
312
2445

$1509 = \bar{x}$
-312
$1197 = \bar{x} - s$
-312
$885 \quad \bar{x} - 2s$

1820 is about the 84th percentile.